BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 6: e91123
doi: 10.3897/biss.6.91123

OPEN ACCESS

Conference Abstract

Open science tools: Supporting hands-on creation of the "digital extended specimen"

Nicky Nicolson‡, Eve Lucas‡

‡ Royal Botanic Gardens, Kew, Richmond, United Kingdom

Corresponding author: Nicky Nicolson (n.nicolson@kew.org)

Received: 01 Aug 2022 | Published: 02 Aug 2022

Citation: Nicolson N, Lucas E (2022) Open science tools: Supporting hands-on creation of the "digital extended specimen". Biodiversity Information Science and Standards 6: e91123. https://doi.org/10.3897/biss.6.91123

Abstract

As a biodiversity informatics community, we have mobilised and interconnected a wide array of information, including specimen collections, published literature and metadata resources, which compile facts about collections and the people that work with them. We have defined data standards to facilitate data interoperability and tools development. Along with colleagues in allied research disciplines, we have helped to develop training resources, enabling researchers to automate routine tasks like data access and reference management. We have also started to explore how we could realise the vision of the digital extended specimen, which would integrate specimens and associated data across multiple research infrastructures, allowing the investigation of wider scale research questions. How this would be achieved is still the subject of discussion and experimentation: an open community will support a diverse range of approaches.

In the construction of the digital extended specimen, we can envision useful activities operating at very different scales: from large scale computational processes run at (or between) research infrastructures, to an ecosystem of lightweight tools that support link construction in context, closer to researchers. A toolset enabling in-context link construction could play a similar role to Open Refine (which has been effective at democratizing data linking between different sources), and supply valuable training data for the development of machine learning approaches.

We will review the use of Open Refine in the biodiversity informatics (and wider research) communities and examine the resources and working practices that facilitated the adoption

© Nicolson N, Lucas E. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

of this tool. We will showcase work towards “an extensible notebook for open science”, and we aim to open a discussion on how a link-aware editor for semi-structured data plus standard open science tools (i.e., those covered by training resources such as software, data and author carpentry) could be viewed as a lightweight alternative to traditional document production—just as Open Refine is a viable alternative to many traditional spreadsheet use cases. Our aim is to enable researchers to develop the digital extended specimen at research time, but without being prescriptive about their workflow.

We will conclude by discussing how this effort supports open science, showing how researchers are *open to access* the data needed to explore their area of study and to form their hypotheses, using well recognised entities (specimens, names, people, institutions, citations etc), represented in data standards and accessed via open APIs - but also *open to choose* how they organise their work.

In the authors own domain (botany), such a tool will be fundamental to e-taxonomic undertakings to build an online reference system in which all known plant species are described, as well as to significant acceleration of parts of the taxonomic process to address the biodiversity crisis.

Keywords

research tools, record linkage, document production

Presenting author

Nicky Nicolson

Presented at

TDWG 2022